图书情报工作
第 62 卷 第 7 期 2018 年 4 月

基于本体的研究主题语义分析方法研究

■ 冯佳 张云秋
吉林大学公共卫生学院 长春 130021

摘要：［**目的/意义**］旨在深入语义层面对研究主题进行分析。［**方法/过程**］提出基于本体的研究主题语义分析方法，从语义类型和与语义关联两个维度展开，并在实证研究过程中，以“医学信息学”为例，对方法进行验证。［**结果/结论**］结果表明，语义类型分析能够辅助研究者对研究主题的内容进行进一步的语义理解；语义关联分析从语义角度分析各个研究主题在语义含义上的关联，在辅助研究者分析某领域研究主题时，能够综合分析相关主题，并发现新的研究交叉点。

关键词：研究主题 语义分析 本体
分类号：G250
DOI:10.13266/j.issn.0252-3116.2018.07.011

1 引言

主题通常是指文章所论述或研究的主要内容。某领域的研究主题能够反映出该领域的研究方向。识别研究主题，把握领域方向，对于科学研究者具有重要意义。近年来，主题模型快速流行，并且被广泛应用于多种语料的主题抽取，如学术语料、Web 本文、社会化媒体资源等。主题模型是对文档中隐含主题的一种建模方法，能够基于文本语料库识别出潜在的主题。目前，主题模型法广泛应用于学术语料，对其进行建模从而识别出研究主题。然而，对于研究主题的分析多依赖研究者的背景知识，且研究者的科研素养和知识背景不同，因而导致分析结果具有较强的主观性。在定量分析方面，对于研究主题的分析可分为基于文献计量学的方法、基于知识图谱的可视化方法。

基于文献计量学的方法大多从研究主题的时间分布、期刊分布、地区分析、国家分布、作者分布等方面结合研究主题的内容进行分析和阐述。如 2014 年，静发冲等人利用文本挖掘的方法，对美国国家科学基金会生物科学部新兴前沿科学处的在研项目进行文本聚类和内容分析，在主题的分析过程中，结合各主题的时间分布和内容信息，展示了各类主题的项目研究内容，并归纳和总结出各类项目的主要特点[1]。

对于基于知识图谱的可视化方法，大多研究者从图谱的节点内容、节点连接强度和节点的位置对研究主题的内容进行分析与解读。如 2009 年栾春娟对 1995-2007 年期间《科学计量学》出版的关于国际专利计量研究的论文和引文进行计量分析，绘制了作者共被引网络、关键词共现网络和作者学术合作群体网络，形象地反映了国际专利计量研究的代表人物和研究主题[2]。2013 年，魏晓峰采用知识图谱对国外信息可视化研究演进、热点主题进行分析，并结合知识图谱进行进一步分析[3]。2014 年，S. Y. Cheng 采用可视化技术对进行电子政府(Electronic Government)领域的研究主题进行识别与分析[4]。

目前，对于研究主题的分析多基于结果的简单呈现，如列表、矩阵或知识图谱等，笔者拟基于本体对研究主题进行映射，借助本体的语义类型和语义结构对研究主题从内容层面进行深入分析。

本体作为一种能在语义和知识层次上描述信息的概念模型建模工具。对于本体的定义，R. Studer 等人给出了较为清晰的解释：“知识本体是对概念体系的明确的、形式化、可共享的规范说明”[5]。“明确”指的是所采用概念的类型及它们应用的约束实行明确的定义，“形式化”指知识本体是能被计算机处理，“共享”是指知识本体应构建相关领域中公认的概念集。通常可以把知识本体看成是“领域知识规范的抽象和描述，表达、共享、重用知识的方法”[6]。

本体构建的初衷是集成某领域的相关知识，提供

作者简介：冯佳(ORCID:0000-0001-9385-3253)，博士研究生；张云秋(ORCID:0000-0002-9790-9581)，教授，博士生导师，通讯作者，E-mail:yunqiu@jlu.edu.cn。
收稿日期:2017-08-23 修回日期:2017-12-27 本文起止页码:96-103 本文责任编辑:王善军

冯佳，张云秋. 基于本体的研究主题语义分析方法研究[J]. 图书情报工作,2018,62(7):96－103.

对该领域知识概念的共同理解,确定领域内共同认可的知识概念。把一个领域的知识抽象成一套概念体系,并以一个个词表来表示,包括每一个词的明确定义、词与词之间的关系以及该领域的一些公理性知识的陈述等,并且能够在该领域的专家之间达成共识,如此便构成了某领域的本体。笔者拟基于领域本体的知识概念结构,并从语义类型和语义关联角度对研究主题进行分析,旨在深入语义层面对其进行分析和解读。

2 研究主题的语义类型分析

词汇或概念的语义类型可以理解为是一种概念属性,可以对概念进行描述和解释。英国语言学家利奇(L. Geoffrey)在《语义学》一书中提出了语义类型这个概念[7]。语义类型是按照语言的运用规律,从语义和人类交际的角度区分的。他将词语的意义分为7种类型:概念意义(Conceptual Meaning)、内涵意义(Connotative Meaning)、社会意义(Social Meaning)、感情意义(Affective Meaning)、联想意义(Reflective Meaning)、搭配意义(Collective Meaning)和主题意义(Thematic Meaning)。语义类型对本体的框架、词元及框架元素进行了抽象与概括,能够表示框架网络中各语义组成部分的固有的、本质的,与词汇所在上下文无关的语义特征。而且和语义类型以一定的逻辑关系构成一个语义类型结构体系,这为本体在自然语言处理中的应用打下了坚实的基础。

目前,语义类型的分析方法主要是语义角色的标注法和基于本体的语义类型分析法。通过语义角色标注法来进行语义类型分析可以对科技文本的研究内容进行系统的分析和解读,提高研究者对科技文本理解的深度和准确度。语义角色标注是对句子中的动词、名词、形容词等进行语义角色标注,通过分析语义角色类型来实现句子级别的浅层语义分析[8]。如2013年,张泽宇等[9]借鉴NCBO Annotator的思想,结合本体知识库和WordNet的语义知识,提出了一种基于语义的文档语义角色标注方法。语义角色标注的重点是对句子中谓词所支配的语义角色(如施事、受事、时间和地点等)进行自动标注。然而在科技文献文本挖掘的过程中,对于科技文献的语义分析,其重点在于分析专业词汇(名称、动词等)的语义类型。

基于本体的语义类型分析是将文本中的词语映射到本体中的概念上,并分析概念的语义类型。本体是一套具有完整结构的概念体系,并在这个体系中,每个概念有其附带的语义类型,可以对概念进行描述和解释,这是通过本体实现语义类型分析的基础。本体是一种概念化的语义表示方法,根据本体思想建立的具有代表性的语义词典有WordNet和HowNet等。有研究者尝试基于本体来分析文本的语义类型,如2007年,张晗等[10]根据UMLS中概念所属的语义类型来挖掘文献间的潜在联系。将本体论应用到语义类型分析,为语义层次上的文本挖掘提供了理论支持。

笔者采用基于本体的语义类型分析,将研究主题的主题词项进行概念映射,将主题词袋转换为“概念词袋”,并深入挖掘词袋中概念的语义类型,使研究主题的分析结果更加丰富。

3 研究主题的语义关联分析

为进一步分析研究主题的语义信息,拟采用语义距离来测度研究主题间的语义相似度。语义相似度能够反映出词汇之间在知识概念和逻辑关系上的关联。笔者拟从语义分析的视角,为研究主题进行语义关联程度的分析,

词语语义上的关系可由领域本体体现。本体是一套概念框架,给出一套词汇来标识一套概念[11]。领域本体包含了领域的概念结构,将概念按照一定的层级结构进行组织。基于领域本体来计算概念语义距离的基础是两个概念具有一定的语义相关性,即概念在本体网络中存在一条通路。术语间概念上的亲疏关系,即术语在本体中的相对位置可以由语义距离来衡量。因此可以借助本体经概念映射,将共现的词汇转换为本体中的术语,并通过本体中术语之间的相对位置来衡量其概念上的亲疏关系。由此,基于领域本体的语义距离能够代表概念在知识上的内在关联程度。

目前针对语义距离的研究相对成熟,科研成果丰富。基于本体的语义距离除了考虑术语间的路径长度外,还考虑到了其他一些因素,如概念层次树的深度、概念层次树的区域密度等。路径长度相同的两个术语,如果位于概念层次的越底层,其语义距离越大;路径长度相同的两个术语,如果位于概念层次树中高密度区域,其语义距离应大于位于低密度区域的。

笔者选取语义距离来度量不同概念在知识上的内在关联程度。语义距离是指概念在本体层次树中的最短路径上每一条边的权值总和[12],通过关联程度的几何度量来有效表征概念间的相似程度。语义距离是衡量两个概念相似度的最基本因素,一般而言它对概念相似度的影响比其他因素都大[13]。

语义距离一般以语义词典为基础,语义词典是对

概念的组织，通常为树状或网状层次结构，多以本体、叙词表等形式表示。以本体为例，这种方法通过概念在本体树中的位置、距离等信息来计算两个概念之间的相似度。两个概念的语义距离与本体树中路径长度与深度有关。常用的语义距离算法有 Leacock Chodorow 法[14]、Weighted Links 法[15]、Wu and Palmer 法[16]等。

笔者选取经典的 Leacock Chodorow 法来计算语义距离，该算法的核心思想是：概念的相似度与概念在本体层次中的路径长度以及本体层次结构的深度有关。计算公式为：

$$\mathrm{Sim}(C_1, C_2) = -\log \frac{\mathrm{len}(C_1, C_2)}{2 \times \mathrm{Depth}} \quad 公式(1)$$

上述公式(1)中 $\mathrm{len}(C_1, C_2)$ 表示概念词 C_1 和 C_2 在本体树中的最短路径长度，Depth 表示本体树的深度。

如何进行研究主题的语义关联分析，本文的研究思路如下：首先进行数据准备，获取研究主题的主题词袋；其次基于领域本体对词袋中的词汇进行概念映射，统计映射后的概念频次；随后截取高频概念构建概念矩阵，并基于领域本体计算概念间的语义距离，最后进行可视化呈现与结果判读，如图 1 所示：

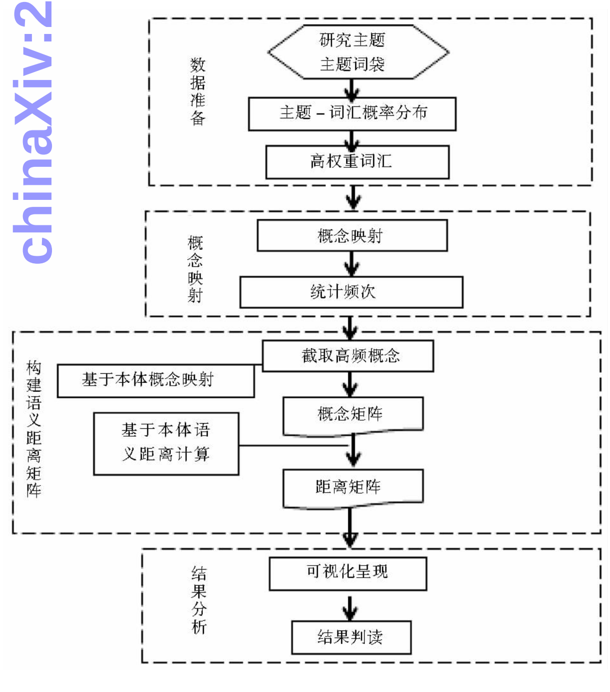


图 1 语义距离计算流程

4 实证分析

医学信息学是一门涉及医学、计算机科学和信息科学的新兴交叉学科。目前，基于医学信息学分析的研究结果不断应用于临床数据分析、药物管理、疾病建模、患者生存预后等方面。因此，准确识别医学信息学领域的研究主题，有助于加强科研管理的战略性导向，对该领域的研究者具有重要的指导意义。笔者以“医学信息学”领域为例，对该领域进行主题抽取，并进行基于本体的主题分析，从而对上述方法进行实证。

4.1 语料库构建

选择 Web of Science 核心合集作为文献集合的数据来源。为全面收集医学信息学领域的相关文献，采用 Web of Science 核心合集的“学科类别”检索功能。Web of Science 核心合集共有 252 个学科类别，“medical informatics”是其中之一。收集“medical informatics”类别下的 2007 至 2016 年的文献，共命中 35 981 的条目(下载时间为 2017 年 1 月 3 日)。随后，基于 LDA 模型抽取该领域的研究主题，共得到医学信息学领域的 19 个主要的研究主题，如表 1 所示：

表 1 医学信息学领域研究主题列表

序号	主题名称	序号	主题名称
1	肿瘤图像分析	11	卫生信息系统评价
2	数据挖掘算法在医学领域的应用	12	疾病生存模型研究
3	医学文本知识提取	13	医学信息学方法与技术研究
4	健康医疗 app	14	电子病历及电子健康记录
5	社区卫生服务研究	15	疾病风险预测
6	临床决策支持研究	16	计算机辅助的疾病诊断
7	基于网络和计算机的新医疗模式	17	机器学习方法在医疗中的应用
8	疾病诊断系统和疾病分类方法研究	18	临床知识语义分析
9	医疗软件的开发与应用	19	大数据背景下医学数据平台构建
10	医疗系统和医疗数据集成研究		

4.2 语义类型分析

基于 LDA 方法识别出的研究主题，采用 MetaMap[17]来实现基于 UMLS 本体映射，将能表征研究主题的主题词转换成 UMLS 本体中的知识概念，使这些主题词的语义得以抽象，将医学信息学领域的研究主题进行概念映射后，部分结果见表 2。

对语义类型进行进一步统计，分析不同主题的语义类型。表 3 列出的是医学信息学领域研究主题的概念及其语义类型。从该领域的语义类型信息来看，主要可以分为以下 7 个方面：

(1)概念类。包括概念实体(Conceptual Entity)、思想或概念(Idea or Concept)、定性概念(Qualitative Concept)、定量概念(Functional Concept)、功能性概念(Functional Concept)、空间概念(Spatial Concept)、时间概念(Temporal Concept)等。

冯佳，张云秋. 基于本体的研究主题语义分析方法研究［J］. 图书情报工作,2018,62(7):96－103.

表 2　主题－概念－语义类型表(部分)

主题	概念	语义类型
T1	Image (Medical Image)	Intellectual Product
	Tumour (Neoplasms)	Neoplastic Process
	Detection	Therapeutic or Preventive Procedure
	Algorithm (algorithm)	Intellectual Product
	MRI (Magnetic Resonance Imaging)	Diagnostic Procedure
	Optimization	Activity
	Shape (Shapes)	Spatial Concept
	3D (Three-dimensional)	Spatial Concept
	CT (Computed Tomography Study File)	Intellectual Product
	Feature (Array Feature)	Conceptual Entity
T2	Algorithm (algorithm)	Intellectual Product
	Performance	Individual Behavior
	Retrieval	Health Care Activity
	Clustering (statistical cluster)	Research Activity
	Optimal (Optimum)	Qualitative Concept
	Language (Languages)	Language
	Query (Question (inquiry))	Intellectual Product
	search (search - EntityNameUse)	Intellectual Product
	Corpus (Body of uterus)	Body Part, Organ, or Organ Component
	Adaptive (adaptive)	Functional Concept
T3	information search (information searching)	Occupational Activity
	Text	Intellectual Product
	Rehabilitation, Medical (Rehabilitation therapy)	Therapeutic or Preventive Procedure
	Biomedical (Biomedicine)	Biomedical Occupation or Discipline
	Clinical	Qualitative Concept
	Old	Temporal Concept
	Semantic (Semantics)	Idea or Conce
T4	Mobile Application (Mobile Applications)	Intellectual Product
	objectives (objective (goal))	Intellectual Product
	Rationale (Indication of (contextual qualifier))	Idea or Concept
	DEVICES (Medical Devices)	Medical Device
	Patients	Patient or Disabled Group
	User (user - Facility type)	Idea or Concept
	Life	Idea or Concept

(2)行为类。包括健康护理活动(Health Care Activity)、活动(Activity)、个人行为(Individual Behavior)、研究活动(Research Activity)、职业活动(Occupational Activity)、教育活动(Educational Activity)、心理过程(Mental Process)、语言(Language)等。

(3)人群类。包括患者和残疾人组(Patient or Disabled Group)、人口(Population Group)、专业和职业组(Professional or Occupational Group)等。

(4)治疗与诊断类。包括治疗或预防程序(Therapeutic or Preventive Procedure)、结果(Finding)、临床属性(Clinical Attribute)、诊断程序(Diagnostic Procedure)。

(5)人体功能与现象类。包括遗传功能(Genetic Function)、生物功能(Organism Function)、肿瘤过程等(Neoplastic Process)。

(6)材料与设备类。包括医疗设备(Medical Device)、制造对象(Manufactured Object)、研究设备(Research Device)等。

(7)职业类。包括职业或学科(Occupation or Discipline)、生物医学职业或学科(Biomedical Occupation or Discipline)。

将语义类型矩阵进行可视化展示,为清晰地呈现结果,图 2 将主题－语义类型的连线是阈值大于等于 2 的显示出来,其中方形节点为研究主题,圆形节点为语义类型,节点间的连线的粗细代表其关联强度。

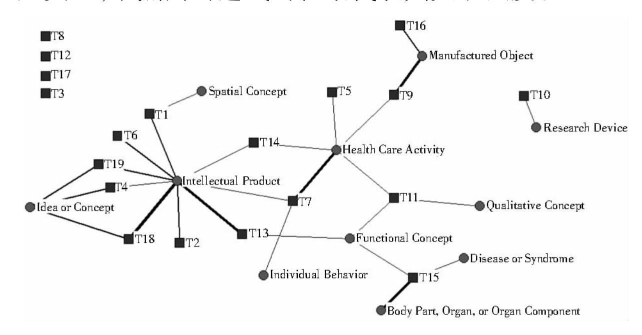


图 2　语义类型图谱

从图 2 中可以看出,医学信息学领域的研究主题具有一些共同的语义类型,如 intellectual product(智力产品)、health care activity(健康医疗活动)、functional concept(功能性概念)等。结合研究主题内容和语义类型可以发现,主题 1(肿瘤图像分析)、主题 2(数据发掘算法的医学应用)、主题 4(健康医疗 app)、主题 6(临床决策支持)、主题 18(临床知识语义分析)、主题 19(大数据背景下的医学数据平台构建)主要围绕着算法、模型、标准、协议、技术等“智力产品”展开研究。

主题 5(社区卫生服务研究)主要针对远程医疗等“健康护理活动”进行研究。主题 7(基于网络和计算机的新医疗模式)涉及医疗干预、自我管理等“健康护理活动”以及新医疗模式下的技术、方法等“智力产品”。主题 15(疾病风险预测)包括心脏、血管等“人体

表 3 语义类型矩阵

语义类型 \ 主题序号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Activity	1					1			1		1								
Biomedical Occupation or Discipline			1																
Body Part, Organ, or Organ Component		1													4				
Classification								1									1		
Clinical Attribute						1						1							
Conceptual Entity	1															1			
Diagnostic Procedure	1																		
Disease or Syndrome															2	1			
Educational Activity								1											
Finding												1				1			1
Functional Concept		1					1		1	1	2		2	1	2	1	1		
Genetic Function										1									1
Health Care Activity		1			2		4		2	1	2		1	2		1		1	
Idea or Conce			1																
Idea or Concept				3		1		1				1	1	1	1		1	3	3
Individual Behavior		1					2	1									1		
Intellectual Product	3	3	1	2	1	3	2	1	1		1	1	4	2		1	1	5	3
Language		1																	
Manufactured Object							1	1	4				1	1		3	1		
Medical Device				1															
Mental Process						1	1										1		
Neoplastic Process	1																		
Nucleic Acid, Nucleoside, or Nucleotide								1									1		
Occupation or Discipline					1					1									
Occupational Activity			1							1									1
Organism Function								1											
Patient or Disabled Group				1										1	1				
Population Group																	1		
Professional or Occupational Group					1	1													
Qualitative Concept		1	1		1			1			2	1					1	1	
Quantitative Concept					1							1							
Research Activity		1										1							1
Research Device										2									
Self-help or Relief Organization					1														
Spatial Concept	2											1				1			
Substance										1									
Temporal Concept			1									1			1				
Therapeutic or Preventive Procedure	1		1																

部位、器官或成分”,冠心病、心衰等“疾病或症状”,分析疾病风险。

综合分析主题的内容和语义类型,有助于为研究主题的解读提供更多信息。

4.3 语义关联分析

对于语义距离的计算,基于 UMLS 本体采用 Leacock Chodorow 法计算概念之间的语义距离,并借助 UMLS::Similarity[18] 在线系统,实现语义距离的计算。为优化可视化效果,选择 z-score[19] 作为标准化方法,得到语义矩阵。

表 4 为相似矩阵,表格中的数值代表不同主题之间的相似性,数值越大相似性越高,语义距离越近。数值为“1”代表同一主题,数值为“0”代表两个主题不具有语义相关度。

表 4　语义矩阵(部分)

主题	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
T1	1	0	0	0	0	0	0	0	0	0
T2	0	1	0.167	0.091 4	0	0.082 6	0	0.1	0	0
T3	0	0.167	1	0.167	0	0.219 7	0	0	0.2	0
T4	0	0.091 4	0.167	1	0	0.068 6	0.1	0	0	0.1
T5	0	0	0	0	1	0.1	0	0	0	0
T6	0	0.082 6	0.219 7	0.068 6	0.1	1	0.1	0	0	0
T7	0	0	0	0.1	0	0.1	1	0	0	0
T8	0	0.1	0	0	0	0	0	1	0	0
T9	0	0	0.2	0	0	0	0	0	1	0
T10	0	0	0	0.1	0	0	0	0	0	1

为进一步分析不同主题间的语义距离，对语义距离矩阵进行可视化，结果如图 3 所示。图 3 中的主题以不同序号和大小显示，圆圈的大小代表主题的相对规模，连线的粗细代表主题之间的语义关联强度。

图 3　语义矩阵可视化图谱

通过对语义关联矩阵可视化图谱的解读，可以辅助研究者综合分析不同的研究主题，例如图 3 中语义关联最强的两个主题是 T17（机器学习方法在医疗中的应用）与 T4（健康医疗 app），结合文档－主题概率分布和主题－词汇概率分布，从语义内容上分析可知健康医疗 app 通过可穿戴设备来监测人体的心电、脑电、肌电信号，并结合深度学习等数据挖掘算法，实现健康数据的分析与管理。

表 5 为研究主题语义矩阵的中心度列表，从表中可以解读出该领域的研究核心和重点，并预测出未来的研究方法。从表 5 可知，在医学信息学领域的研究主题中，中心度最高的为 T4（健康医疗 app），这表明目前该领域的研究重点在此，并且健康医疗 app 的研究涉及医学信息学领域的各个研究方向，如 T17（机器学习方法在医疗中的应用）、T2（数据挖掘算法在医学领域的应用）、T13（医学信息学方法与技术研究）、T9（医疗软件的开发与应用）等多个方面，健康医疗 app 需要综合医学信息学领域的多种方法和技术，因此针对健康医疗 app 的研发过程中涉及到的方法、技术和软件开发与应用是医学信息学领域未来的研究方向。

表 5　研究主题语义矩阵中心度列表

主题	绝对点度中心度 Degree	相对点度中心度 NrmDegree	份额 Share
T8：健康医疗 app	1.49	16.605	0.126
T53：机器学习方法在医疗中的应用	1.467	16.353	0.124
T17：临床决策支持研究	1.132	12.615	0.096
T31：卫生信息系统评价	0.9	10.031	0.076
T3：数据挖掘算法在医学领域的应用	0.843	9.39	0.071
T41：医学信息学方法与技术研究	0.8	8.916	0.068
T24：医疗系统和医疗数据集成研究	0.8	8.916	0.068
T9：社区卫生服务研究	0.774	8.626	0.065
T46：疾病风险预测	0.718	7.997	0.061
T58：大数据背景下医学数据平台构建研究	0.6	6.687	0.051
T45：电子病历及电子健康记录	0.5	5.573	0.042
T22：医疗软件的开发与应用	0.4	4.458	0.034
T34：疾病生存模型研究	0.3	3.344	0.025
T19：基于网络和计算机的新医疗模式	0.2	2.229	0.017
T21：疾病诊断系统和疾病分类方法研究	0.2	2.229	0.017
T4：医学文本知识提取	0.2	2.229	0.017
T48：计算机辅助的疾病诊断	0.2	2.229	0.017
T54：临床知识语义分析	0.2	2.229	0.017
T1：肿瘤图像分析	0	0	0

从语义关联图谱中,可以看出医学信息学领域各个主题的语义相关信息,在分析研究主题的过程中,可以发现某个研究主题与该领域的哪些研究主题在语义内容上具有一定的关联。这种方法在一定程度上体现出了专家的智慧,是一种人工智能化的结果分析,可以使医学信息学领域的研究者快速、直观、清晰地理解本领域的研究现状;对于领域的研究专家,该结果能够辅助其对领域信息的研判,并在主题关联的分析中,发现新的、具有价值的交叉研究点。

5 结论

笔者提出基于本体的研究主题语义分析方法,从语义类型分析和语义关联分析两个维度展开,语义类型分析能够辅助研究者对研究主题的内容进行进一步的语义理解;语义关联分析从语义角度分析了各个研究主题在语义含义上的关联,能够辅助研究者在分析某领域研究主题时,综合分析相关主题,并发现新的研究交叉点。

本文在研究主题分析方面进行了一定的探索,然而仅针对研究主题的语义类型和语义关联进行了简明的分析。在未来的研究工作中,将进一步挖掘研究主题的语义信息,结合目前成熟的语义分析技术,对科技文献中的研究主题进行更加深入的探索,以期为科技创新和科技决策提供支持和帮助。

参考文献:

[1] 静发冲,李晨英,韩明杰,等.基于文本挖掘的美国 NSF 生物科学部新兴前沿项目主题分析[J].现代情报,2014,34(12):107-112.
[2] 栾春娟, 王续琨, 刘则渊, 等. 专利计量研究国际前沿的计量分析[J].科学学研究,2008,26(2):334-338,310.
[3] 魏晓峰. 基于知识图谱的国外信息可视化研究演进、热点与前沿分析[J].情报学报,2013,32(5):533-547.
[4] CHENG S Y, DING L. Mapping of electronic government: the trend of research fronts[C]//2014 seventh international joint conference on computational sciences and optimization. Piscataway: IEEE, 2014:509-513. doi:10.1109/CSO.2014.100.
[5] STUDER R, BENJAMINS V R, FENSEL D. Knowledge engineering principles and methods[J]. Data and knowledge engineering, 1998,25(1/2):161-197.
[6] 刘炜,李大玲,夏翠娟.元数据与知识本体[J].图书馆杂志,2004,23(6):50-54,49.
[7] LEECH G. 语义学[M]. 李瑞华,王彤福,杨自俭,等译.上海:上海外语教育出版社,1987.
[8] 杨选选,张蕾. 基于语义角色和概念图的信息抽取模型[J].计算机应用,2010, 30(2):411-414.
[9] 张泽宇,李莉,谭凤, 等. 基于语义的文档标注方法研究[J].计算机工程与科学, 2013,35(9): 151-156.
[10] 张晗, 任志国, 于倩, 等. 基于 UMLS 医学本体的挖掘文献间潜在联系的设计与实现[J]. 现代图书情报技术, 2007, 2(9): 72-75.
[11] BOINO D, CORNO F, PESCARMONA F. Automatic learning of text-to-concept mappings exploiting WordNet-like lexical networks [C]//Proceedings of the 2005 ACM symposium on Applied computing. New York: ACM, 2005: 1639 - 1644. doi: 10. 1145/1066677.1067050.
[12] 徐德智,邓春卉. PASSI K. 基于SUMO 的概念语义相似度研究[J].计算机应用,2006, 26(1): 180-183.
[13] 唐中林.基于本体的概念相似度计算方法的研究[D]. 武汉:武汉理工大学, 2013.
[14] LEACOCK C, CHODOROW M. Combining local context and WordNet similarity for word sense identification[M]. Massachusetts: MIT Press,1998:265-283.
[15] RICHARDSON R, SMEATON A F. Using WordNet in a knowledge-based approach to information retrieval[EB/OL].[2017-06-15]. http://citeseerx. ist. psu. edu/viewdoc/download; jsessionid=C386BEFF23C88F1BAA73F65BD484FD08? doi=10.1.1.48.9324&rep=rep1&type=pdf.
[16] WU Z, PALMER M. Verbs semantics and lexical selection[C]//Proceedings of the 32nd annual meeting on Association for Computational Linguistics. Stroudsburg, PA:Association for Computational Linguistics, 1994: 133-138. doi:10.3115/981732.981751.
[17] U. S. National Library of Medicine. Interactive MetaMap[EB/OL]. [2017-06-15]. http://ii. nlm. nih. gov/Interactive/UTS_Required/metamap. shtml.
[18] PEDERSEN T. UMLS: Similarity[EB/OL]. [2017-06-15]. http://atlas. ahc. umn. edu/cgi-bin/umls_similarity. cgi.
[19] CFOO L, MAFAUZY M. Does the use of mean or median Z-score of the thyroid volume indices provide a more precise description of the iodine deficiency disorder status of a population? [J]. European journal of endocrinology, 1999, 141(6): 557-560.

作者贡献说明:

冯佳:提出研究思路和论文框架,进行实验并收集数据,撰写论文并修改;

张云秋:确定论文选题,提出修改意见,进行研究内容的完善和修改。

冯佳，张云秋. 基于本体的研究主题语义分析方法研究[J]. 图书情报工作,2018,62(7):96－103.

Study on Semantic Analysis Method of Research Topics Based on Ontology

Feng Jia　Zhang Yunqiu
Public Health School, Jilin University, Changchun 130021

Abstract: [**Purpose/significance**] This paper aims at analyzing the research topics by going deeper into the semantic dimension. [**Method/process**] This paper proposed a semantic analysis method based on ontology, which includes the semantic type analysis and semantic relevance analysis. Then, in the empirical study, this paper took "medical informatics" as an example to verify the method. [**Result/conclusion**] This paper reveals that semantic type analysis can help researchers make a further semantic understanding for the research topics. Semantic relevance analysis analyze the semantic meaning of each research topic from the semantic perspective, when assisting researchers in analyzing a research topic in a field, it can realize the relevance analysis of every topic synthetically, and find some research intersections.

Keywords: research topic　semantic analysis　ontology